



Red Hat Enterprise Linux Release 9.2 Manual Pages on 'diff.1p' command

\$ man diff.1p

DIFF(1P) POSIX Programmer's Manual DIFF(1P)

PROLOG

This manual page is part of the POSIX Programmer's Manual. The Linux implementation of this interface may differ (consult the corresponding Linux manual page for details of Linux behavior), or the interface may not be implemented on Linux.

NAME

diff ? compare two files

SYNOPSIS

```
diff [-c|-e|-f|-u|-C n|-U n] [-br] file1 file2
```

DESCRIPTION

The diff utility shall compare the contents of file1 and file2 and write to standard output a list of changes necessary to convert file1 into file2. This list should be minimal. No output shall be produced if the files are identical.

OPTIONS

The diff utility shall conform to the Base Definitions volume of POSIX.1?2017, Section 12.2, Utility Syntax Guidelines.

The following options shall be supported:

- b Cause any amount of white space at the end of a line to be treated as a single <newline> (that is, the white-space characters preceding the <newline> are ignored) and other strings of white-space characters, not including <newline> charac?

ters, to compare equal.

- c Produce output in a form that provides three lines of copied context.
- C n Produce output in a form that provides n lines of copied context (where n shall be interpreted as a positive decimal integer).
- e Produce output in a form suitable as input for the ed utility, which can then be used to convert file1 into file2.
- f Produce output in an alternative form, similar in format to -e, but not intended to be suitable as input for the ed utility, and in the opposite order.
- r Apply diff recursively to files and directories of the same name when file1 and file2 are both directories.
The diff utility shall detect infinite loops; that is, entering a previously visited directory that is an ancestor of the last file encountered. When it detects an infinite loop, diff shall write a diagnostic message to standard error and shall either recover its position in the hierarchy or terminate.
- u Produce output in a form that provides three lines of unified context.
- U n Produce output in a form that provides n lines of unified context (where n shall be interpreted as a non-negative decimal integer).

OPERANDS

The following operands shall be supported:

file1, file2

A pathname of a file to be compared. If either the file1 or file2 operand is '-', the standard input shall be used in its place.

If both file1 and file2 are directories, diff shall not compare block special files, character special files, or FIFO special files to any files and shall not compare regular files to directories. Further de?

tails are as specified in Diff Directory Comparison Format. The behavior of diff on other file types is implementation-defined when found in directories.

If only one of file1 and file2 is a directory, diff shall be applied to the non-directory file and the file contained in the directory file with a filename that is the same as the last component of the non-directory file.

STDIN

The standard input shall be used only if one of the file1 or file2 operands references standard input. See the INPUT FILES section.

INPUT FILES

The input files may be of any type.

ENVIRONMENT VARIABLES

The following environment variables shall affect the execution of diff:

LANG Provide a default value for the internationalization variables that are unset or null. (See the Base Definitions volume of POSIX.1:2017, Section 8.2, Internationalization Variables for the precedence of internationalization variables used to determine the values of locale categories.)

LC_ALL If set to a non-empty string value, override the values of all the other internationalization variables.

LC_CTYPE Determine the locale for the interpretation of sequences of bytes of text data as characters (for example, single-byte as opposed to multi-byte characters in arguments and input files).

LC_MESSAGES

Determine the locale that should be used to affect the format and contents of diagnostic messages written to standard error and informative messages written to standard output.

LC_TIME Determine the locale for affecting the format of file time stamps written with the -C and -c options.

NLSPATH Determine the location of message catalogs for the processing of LC_MESSAGES.

TZ Determine the timezone used for calculating file timestamps written with a context format. If TZ is unset or null, an unspecified default timezone shall be used.

ASYNCHRONOUS EVENTS

Default.

STDOUT

Diff Directory Comparison Format

If both file1 and file2 are directories, the following output formats shall be used.

In the POSIX locale, each file that is present in only one directory shall be reported using the following format:

"Only in %s: %s\n", <directory pathname>, <filename>

In the POSIX locale, subdirectories that are common to the two directories may be reported with the following format:

"Common subdirectories: %s and %s\n", <directory1 pathname>, <directory2 pathname>

For each file common to the two directories, if the two files are not to be compared: if the two files have the same device ID and file serial number, or are both block special files that refer to the same device, or are both character special files that refer to the same device, in the POSIX locale the output format is unspecified. Otherwise, in the POSIX locale an unspecified format shall be used that contains the pathnames of the two files.

For each file common to the two directories, if the files are compared and are identical, no output shall be written. If the two files differ, the following format is written:

"diff %s %s %s\n", <diff_options>, <filename1>, <filename2>

where <diff_options> are the options as specified on the command line.

All directory pathnames listed in this section shall be relative to the original command line arguments. All other names of files listed in this section shall be filenames (pathname components).

Diff Binary Output Format

In the POSIX locale, if one or both of the files being compared are not

text files, it is implementation-defined whether diff uses the binary file output format or the other formats as specified below. The binary file output format shall contain the pathnames of two files being compared and the string "differ".

If both files being compared are text files, depending on the options specified, one of the following formats shall be used to write the differences.

Diff Default Output Format

The default (without -e, -f, -c, -C, -u, or -U options) diff utility output shall contain lines of these forms:

```
"%da%d\n", <num1>, <num2>
```

```
"%da%d,%d\n", <num1>, <num2>, <num3>
```

```
"%dd%d\n", <num1>, <num2>
```

```
"%d,%dd%d\n", <num1>, <num2>, <num3>
```

```
"%dc%d\n", <num1>, <num2>
```

```
"%d,%dc%d\n", <num1>, <num2>, <num3>
```

```
"%dc%d,%d\n", <num1>, <num2>, <num3>
```

```
"%d,%dc%d,%d\n", <num1>, <num2>, <num3>, <num4>
```

These lines resemble ed subcommands to convert file1 into file2. The line numbers before the action letters shall pertain to file1; those after shall pertain to file2. Thus, by exchanging a for d and reading the line in reverse order, one can also determine how to convert file2 into file1. As in ed, identical pairs (where num1= num2) are abbreviated as a single number.

Following each of these lines, diff shall write to standard output all lines affected in the first file using the format:

```
"< %s", <line>
```

and all lines affected in the second file using the format:

```
"> %s", <line>
```

If there are lines affected in both file1 and file2 (as with the c subcommand), the changes are separated with a line consisting of three <hyphen-minus> characters:

```
"---\n"
```

Diff -e Output Format

With the `-e` option, a script shall be produced that shall, when provided as input to `ed`, along with an appended `w` (write) command, convert `file1` into `file2`. Only the `a` (append), `c` (change), `d` (delete), `i` (insert), and `s` (substitute) commands of `ed` shall be used in this script. Text lines, except those consisting of the single character `<period>` (`.`), shall be output as they appear in the file.

Diff -f Output Format

With the `-f` option, an alternative format of script shall be produced.

It is similar to that produced by `-e`, with the following differences:

1. It is expressed in reverse sequence; the output of `-e` orders changes from the end of the file to the beginning; the `-f` from beginning to end.
2. The command form `<lines> <command-letter>` used by `-e` is reversed. For example, `10c` with `-e` would be `c10` with `-f`.
3. The form used for ranges of line numbers is `<space>-separated`, rather than `<comma>-separated`.

Diff -c or -C Output Format

With the `-c` or `-C` option, the output format shall consist of affected lines along with surrounding lines of context. The affected lines shall show which ones need to be deleted or changed in `file1`, and those added from `file2`. With the `-c` option, three lines of context, if available, shall be written before and after the affected lines. With the `-C` option, the user can specify how many lines of context are written. The exact format follows.

The name and last modification time of each file shall be output in the following format:

```
**** %s %s\n", file1, <file1 timestamp>
```

```
--- %s %s\n", file2, <file2 timestamp>
```

Each `<file>` field shall be the pathname of the corresponding file being compared. The pathname written for standard input is unspecified.

In the POSIX locale, each `<timestamp>` field shall be equivalent to the output from the following command:

date "+%a %b %e %T %Y"

without the trailing <newline>, executed at the time of last modification of the corresponding file (or the current time, if the file is standard input).

Then, the following output formats shall be applied for every set of changes.

First, a line shall be written in the following format:

"*****\n"

Next, the range of lines in file1 shall be written in the following format if the range contains two or more lines:

"*** %d,%d ***\n", <beginning line number>, <ending line number>

and the following format otherwise:

"*** %d ***\n", <ending line number>

The ending line number of an empty range shall be the number of the preceding line, or 0 if the range is at the start of the file.

Next, the affected lines along with lines of context (unaffected lines) shall be written. Unaffected lines shall be written in the following format:

" %s", <unaffected_line>

Deleted lines shall be written as:

"- %s", <deleted_line>

Changed lines shall be written as:

"! %s", <changed_line>

Next, the range of lines in file2 shall be written in the following format if the range contains two or more lines:

"--- %d,%d ---\n", <beginning line number>, <ending line number>

and the following format otherwise:

"--- %d ----\n", <ending line number>

Then, lines of context and changed lines shall be written as described in the previous formats. Lines added from file2 shall be written in the following format:

"+ %s", <added_line>

The `-u` or `-U` options behave like the `-c` or `-C` options, except that the context lines are not repeated; instead, the context, deleted, and added lines are shown together, interleaved. The exact format follows.

The name and last modification time of each file shall be output in the following format:

```
"--- %s\t%s%s %s\n", file1, <file1 timestamp>, <file1 frac>, <file1 zone>
```

```
"+++ %s\t%s%s %s\n", file2, <file2 timestamp>, <file2 frac>, <file2 zone>
```

Each `<file>` field shall be the pathname of the corresponding file being compared, or the single character `'-'` if standard input is being compared. However, if the pathname contains a `<tab>` or a `<newline>`, or if it does not consist entirely of characters taken from the portable character set, the behavior is implementation-defined.

Each `<timestamp>` field shall be equivalent to the output from the following command:

```
date '+%Y-%m-%d %H:%M:%S'
```

without the trailing `<newline>`, executed at the time of last modification of the corresponding file (or the current time, if the file is standard input).

Each `<frac>` field shall be either empty, or a decimal point followed by at least one decimal digit, indicating the fractional-seconds part (if any) of the file timestamp. The number of fractional digits shall be at least the number needed to represent the file's timestamp without loss of information.

Each `<zone>` field shall be of the form `"shhmm"`, where `"shh"` is a signed two-digit decimal number in the range `-24` through `+25`, and `"mm"` is an unsigned two-digit decimal number in the range `00` through `59`. It represents the timezone of the timestamp as the number of hours (`hh`) and minutes (`mm`) east (+) or west (-) of UTC for the timestamp. If the hours and minutes are both zero, the sign shall be `'+'`. However, if the timezone is not an integral number of minutes away from UTC, the `<zone>` field is implementation-defined.

Then, the following output formats shall be applied for every set of changes.

First, the range of lines in each file shall be written in the following

format:

```
"@@ -%s +%s @@", <file1 range>, <file2 range>
```

Each <range> field shall be of the form:

```
"%1d", <beginning line number>
```

or:

```
"%1d,1", <beginning line number>
```

if the range contains exactly one line, and:

```
"%1d,%1d", <beginning line number>, <number of lines>
```

otherwise. If a range is empty, its beginning line number shall be the number of the line just before the range, or 0 if the empty range starts the file.

Next, the affected lines along with lines of context shall be written.

Each non-empty unaffected line shall be written in the following format:

mat:

```
" %s", <unaffected_line>
```

where the contents of the unaffected line shall be taken from file1.

It is implementation-defined whether an empty unaffected line is written as an empty line or a line containing a single <space> character.

This line also represents the same line of file2, even though file2's line may contain different contents due to the -b. Deleted lines shall be written as:

```
"-%s", <deleted_line>
```

Added lines shall be written as:

```
"+%s", <added_line>
```

The order of lines written shall be the same as that of the corresponding file. A deleted line shall never be written immediately after an added line.

If -U n is specified, the output shall contain no more than 2n consecutive unaffected lines; and if the output contains an affected line and this line is adjacent to up to n consecutive unaffected lines in the corresponding file, the output shall contain these unaffected lines.

-u shall act like -U3.

STDERR

The standard error shall be used only for diagnostic messages.

OUTPUT FILES

None.

EXTENDED DESCRIPTION

None.

EXIT STATUS

The following exit values shall be returned:

0 No differences were found.

1 Differences were found.

>1 An error occurred.

CONSEQUENCES OF ERRORS

Default.

The following sections are informative.

APPLICATION USAGE

If lines at the end of a file are changed and other lines are added, diff output may show this as a delete and add, as a change, or as a change and add; diff is not expected to know which happened and users should not care about the difference in output as long as it clearly shows the differences between the files.

EXAMPLES

If dir1 is a directory containing a directory named x, dir2 is a directory containing a directory named x, dir1/x and dir2/x both contain files named date.out, and dir2/x contains a file named y, the command:

```
diff -r dir1 dir2
```

could produce output similar to:

```
Common subdirectories: dir1/x and dir2/x
```

```
Only in dir2/x: y
```

```
diff -r dir1/x/date.out dir2/x/date.out
```

```
1c1
```

```
< Mon Jul 2 13:12:16 PDT 1990
```

```
---
```

```
> Tue Jun 19 21:41:39 PDT 1990
```

RATIONALE

The `-h` option was omitted because it was insufficiently specified and does not add to applications portability.

Historical implementations employ algorithms that do not always produce a minimum list of differences; the current language about making every effort is the best this volume of POSIX.1?2017 can do, as there is no metric that could be employed to judge the quality of implementations against any and all file contents. The statement "This list should be minimal" clearly implies that implementations are not expected to provide the following output when comparing two 100-line files that differ in only one character on a single line:

```
1,100c1,100
```

```
all 100 lines from file1 preceded with "< "
```

```
---
```

```
all 100 lines from file2 preceded with "> "
```

The "Only in" messages required when the `-r` option is specified are not used by most historical implementations if the `-e` option is also specified. It is required here because it provides useful information that must be provided to update a target directory hierarchy to match a source hierarchy. The "Common subdirectories" messages are written by System V and 4.3 BSD when the `-r` option is specified. They are allowed here but are not required because they are reporting on something that is the same, not reporting a difference, and are not needed to update a target hierarchy.

The `-c` option, which writes output in a format using lines of context, has been included. The format is useful for a variety of reasons, among them being much improved readability and the ability to understand difference changes when the target file has line numbers that differ from another similar, but slightly different, copy. The patch utility is most valuable when working with difference listings using a context format. The BSD version of `-c` takes an optional argument specifying the amount of context. Rather than overloading `-c` and breaking the Utility Syntax Guidelines for `diff`, the standard developers decided to add a

separate option for specifying a context diff with a specified amount of context (-C). Also, the format for context diffs was extended slightly in 4.3 BSD to allow multiple changes that are within context lines from each other to be merged together. The output format contains an additional four <asterisk> characters after the range of affected lines in the first filename. This was to provide a flag for old programs (like old versions of patch) that only understand the old context format. The version of context described here does not require that multiple changes within context lines be merged, but it does not prohibit it either. The extension is upwards-compatible, so any vendors that wish to retain the old version of diff can do so by adding the extra four <asterisk> characters (that is, utilities that currently use diff and understand the new merged format will also understand the old unmerged format, but not vice versa).

The -u and -U options of GNU diff have been included. Their output format, designed by Wayne Davison, takes up less space than -c and -C format, and in many cases is easier to read. The format's timestamps do not vary by locale, so LC_TIME does not affect it. The format's line numbers are rendered with the %1d format, not %d, because the file format notation rules would allow extra <blank> characters to appear around the numbers.

The substitute command was added as an additional format for the -e option. This was added to provide implementations with a way to fix the classic ``dot alone on a line'' bug present in many versions of diff.

Since many implementations have fixed this bug, the standard developers decided not to standardize broken behavior, but rather to provide the necessary tool for fixing the bug. One way to fix this bug is to output two periods whenever a lone period is needed, then terminate the append command with a period, and then use the substitute command to convert the two periods into one period.

The BSD-derived -r option was added to provide a mechanism for using diff to compare two file system trees. This behavior is useful, is standard practice on all BSD-derived systems, and is not easily repro-

ducible with the find utility.

The requirement that diff not compare files in some circumstances, even though they have the same name, is based on the actual output of historical implementations. The specified behavior precludes the problems arising from running into FIFOs and other files that would cause diff to hang waiting for input with no indication to the user that diff was hung. An earlier version of this standard specified the output format more precisely, but in practice this requirement was widely ignored and the benefit of standardization seemed small, so it is now unspecified. In most common usage, diff -r should indicate differences in the file hierarchies, not the difference of contents of devices pointed to by the hierarchies.

Many early implementations of diff require seekable files. Since the System Interfaces volume of POSIX.1-2017 supports named pipes, the standard developers decided that such a restriction was unreasonable.

Note also that the allowed filename - almost always refers to a pipe.

No directory search order is specified for diff. The historical order?

ing is, in fact, not optimal, in that it prints out all of the differ?

ences at the current level, including the statements about all common subdirectories before recursing into those subdirectories.

The message:

```
"diff %s %s %s\n", <diff_options>, <filename1>, <filename2>
```

does not vary by locale because it is the representation of a command, not an English sentence.

FUTURE DIRECTIONS

None.

SEE ALSO

cmp, comm, ed, find

The Base Definitions volume of POSIX.1-2017, Chapter 8, Environment Variables, Section 12.2, Utility Syntax Guidelines

COPYRIGHT

Portions of this text are reprinted and reproduced in electronic form from IEEE Std 1003.1-2017, Standard for Information Technology -- Por?

table Operating System Interface (POSIX), The Open Group Base Specifications Issue 7, 2018 Edition, Copyright (C) 2018 by the Institute of Electrical and Electronics Engineers, Inc and The Open Group. In the event of any discrepancy between this version and the original IEEE and The Open Group Standard, the original IEEE and The Open Group Standard is the referee document. The original Standard can be obtained online at <http://www.opengroup.org/unix/online.html> .

Any typographical or formatting errors that appear in this page are most likely to have been introduced during the conversion of the source files to man page format. To report such errors, see https://www.kernel.org/doc/man-pages/reporting_bugs.html .

IEEE/The Open Group

2017

DIFF(1P)